FUZZY LOGIC-BASED ASSESSMENT ADJUSTS OF MULTIPLE CHOICE QUESTIONNAIRES

J. Domingo-Peña
Department of Automatic Control (ESAII), Eastern Barcelona School of Engineering (EEBE), Technical University of Catalonia – BarcelonaTech (UPC)
Barcelona, Spain
0000-0001-5776-4720

H. Martínez-García [1]
Department of Electronic Engineering, Eastern Barcelona School of Engineering (EEBE), Technical University of Catalonia – BarcelonaTech (UPC)
Barcelona, Spain
0000-0002-7977-2577, and 0000-0002-1446-0652

Conference Key Areas: *Artificial intelligence in education, Mathematics at the heart of Engineering.*
Keywords: *Fuzzy logic, objective tests, multiple-choice questionnaires, grade correction, students' assessment.*

ABSTRACT

In University environment, it is common to use multiple-choice objective tests with three or four possible answers, of which only one is correct and the rest are erroneous. In this type of tests, usually the wrong answers are penalized in order to avoid the effect of the random answers. However, there are questions that hardly students answer since their difficulty is high. On the other hand, there are also questions that answer virtually all students since their difficulty is simple. While sometimes the course professor chooses to suppress these questions, it is also common to leave them as part of the calculation of the overall score. This communication proposes a way of, without suppressing any question, making a readjustment of the grades based on fuzzy logic techniques. To do this, it is considered, on the one hand, the initial grade obtained by each student and, on the other, the total difficulty index of the test. With these two variables, an approximation can be made to a system of linguistic variables that allows correcting the final grades of each student based on the objective difficulty of the test and a set of rules established by the professor. This will revert to greater "justice" in students' mark system, since it will be a function of the difficulty of the test.

[1] *Corresponding Author*
H. Martínez-García
herminio.martinez@upc.edu

1 INTRODUCTION

1.1 Assessment system

Student assessment is the process that allows knowing what is the level achieved by these students in certain formative objectives. A fair and transparent evaluation assessment system is the one that guarantees that the student body receives a grade that does not limit their future opportunities. Notice that to approve or fair unfairly is vital and the qualification must be, first of all, fair for each student, and as close as possible to the reality of the knowledge achieved for him/her, since each person is unique and, therefore, the assessment should take into account this diversity.

In order to assess the students' performance, in terms of knowledge acquired, one of the most widely used tests is the multiple choice questionnaire (MCQ), consisting of a question and three or four possible answers. Only one of them is valid and the others, called distractors, are false. These formats are known, in a short form, as MCQ-3 and MCQ-4, respectively.

Research of [1], [2] and [3] suggest the model with three answers (MCQ-3) since it saves response time to the student and does not oblige the professor, especially junior, to use improbable and/or defective distractors. On the other hand, the four-answer model is more suitable if there is a reasonable time available for the test or if the professor is already a senior.

On the other hand, since Zadeh [4] established the principles of fuzzy logic, this has been used to solve problems in very different fields, calling attention also in the academic field as it helps to obtain fairer grades in students and with greater transparency. For instance, in a course/subject rated over 100 points, in which at least 50 points are needed to pass it, the case of a student with a grade of 49 would be considered as a "questionable" mark since it could be considered approved or not, taking into account factors such as the time taken to answer the test, its difficulty, the importance of each question, the complexity in determining the correct answer, etc. Therefore, these factors should be considered in some way when course professor defined the mark obtained by his/her student. In addition, the teacher may have different types of students in his classroom with different learning rhythms or particular educational needs, which means that not all of them, perhaps, should be evaluated in the same way, since the classroom is a very diverse human space. With fuzzy adjusts it's possible be a more accurate assessment. With diffuse adjustments, it is possible that the evaluation is more precise and even fairer by being able to attend to particular cases.

1.2 Literature Review Concerning Recalculation of Students' Assessment Using Fuzzy Logic

Following [5], fuzzy logic is a generalization of crisp set theory and traditional dual-valued logic. In fact, in fuzzy set theory the range of the membership function characterizing a set is extended from *only two values*, 0 or 1, to *any* value between 0 and 1.

Most of publications related to student assessment with fuzzy logic techniques refer to seminal articles of Zadeh and Biswas [6]. The article by Chen and Lee [7], also related to students' assessment through fuzzy logic, is also cited profusely in the literature. Since then, fuzzy logic has been used in very different academic fields, being a current of interest the adjustment of the grades obtained by students taking different adjustment parameters and mixing fuzzy logic with other techniques traditionally associated with artificial intelligence (AI) as the case of data mining. One interesting example is in [8], [9] and [10].

They are in the literature many different approachs using fuzzy logic, as can see in [11] and [12], and also in [13] to [18]; in many approaches with multiple input variables, it seems reasonable to use a fuzzy inference model by successive steps, since it allows establishing the rule bases by pairs of variables, giving rise to small and simple base rules simply deductible by the experience.

Studies carried out related to the different fuzzy inference techniques should also be pointed out, and defuzzification methods like those of related in [19], [20] and [21]. In addition, many authors use Matlab® to implement their contributions.

Regarding the geometry of the fuzzy sets used, almost all the publications consulted refer to triangular or trapezoidal sets due to the simplicity of calculation since the differences that would be obtained with other joint geometries (sigmoidal –used by [12]–; Gaussian –used by Hameed and Sorensen [17]–, hyperbolic tangents, etc.) are not relevant to the outcome. Also, the use of other geometries that are not trapezoidal or rectangular are used only when the use of Matlab® software is available.

We propose in the following section a very simple method of calculation using the inference of Mamdani [22], and taking into consideration only the difficulty of the test and the qualification obtained. Some authors such [19] have used the difficulty of a certain competence based on students' marks who studied the subject before and with a difficulty level specified by the professor. In any case, the use of the difficulty index of an objective test has not been located in the bibliography consulted by authors.

2 METHODOLOGY

In an MCQ, the calculation of the difficulty index (*DI*) of each of the questions is the one expressed in (1):

$$DI = \frac{\#Correct\ answers}{\#Questions} \tag{1}$$

It is considered a classic classification of the difficulty scale, taken from [23], which is the one that can be seen in Table 1. The scale and difficulty may vary, but authors consider that that listed in Table 1 is very convenient for academic purposes.

The overall difficulty index of an entire test, DI_{TOTAL}, may be established by taking the average of all difficulty indexes for all questions, according to (2). Therefore, it is simple, by means of a simple table with a spreadsheet that contains in the rows to the students and in the columns the questions, to have in each cell the qualification obtained by each student in each question. Applying (1) to each column, you can determine the parameter DI and identify the difficulty of each question. In addition, applying (2) the difficulty of the test, as a whole, can be obtained.

Table 1. Classic interpretation of the difficulty index (DI).

Value	***Interpretation***	***Suggested Action for the question***
DI > 0.75	Very Easy	Dismiss definitively
0.56 < DI < =0.75	Easy	Candidate to be discarded
0.46 < DI < =0.55	Regular	Review
0.26 < DI < =0.45	Difficult	Candidate to be discarded
DI <= 0.25	Very Difficult	Dismiss definitively

For MCQ-3, the optimal total difficulty would be 0.67 [24]. However, it is considered average difficulty, in general, between 0.50 and 0.60 [25]. If possible, experts suggest that items should have indices of difficulty no less than 0.20 and no greater than 0.80. It is desirable to have most items in the 0.30 to 0.50 range of difficulty. Very difficult and very easy items contribute little to the discriminating power of a test [26].

$$DI_{TOTAL} = \frac{\text{Mean of Difficulty Indexes}}{\text{Number of Indexes}} \tag{2}$$

When differentiating one category of difficulty from the next using crisp values, it is difficult to consider that a question with DI = 0.55 is “regular” and that one with DI = 0.56 is “easy” so that it seems natural to consider this variable as a fuzzy set taking as many subsets as difficulty categories. Likewise, when considering the DI_{TOTAL}, the overall difficulty of the test can be established diffusely, considering that a test with DI_{TOTAL} = 0.55 is “regular” and one with DI_{TOTAL} = 0.56 is “easy”. For this reason, and taking as reference the crisp values of this indicator, a categorization based in fuzzy subsets is established such as the proposed in Figure 1. In this article, we use triangular membership functions because this shape has been proven popular in fuzzy logic and being used extensively in student academic performance assessment [26].

In Figure 1 it can be observed, respecting the spirit of the differentiation of difficulty index, that values below 0.25 and above 0.75 as well as those between 0.45 and 0.55, will only belong to one fuzzy subset, while that the rest of values will belong to two fuzzy subsets.

Taking as reference the crisp values of this indicator, a categorization based in fuzzy subsets is established such as the proposed in Figure 1. We use triangular membership functions [27], but other shapes are possible (that not has a significative difference respect to results).

Let us take as an example, a test with 5 questions in format type MCQ-3 (three answers per question, an answer right and two answers acting as distractors), in which five students participated. In order to discard the random factor and, in some way, discourage the student to try his luck in case he does not clearly know the answer. The amount of penalty would be established, in this case, by the expression (3) taken from [28] which, in turn, [29] cites and also is studied by [30] and [31]. The unanswered questions will have zero value and the wrong answers, will have the corresponding penalty according to their weight.

$$Penalty = \frac{1}{\#Possible\ answers - 1} = \frac{1}{3-1} = 0{,}5 \quad (3)$$

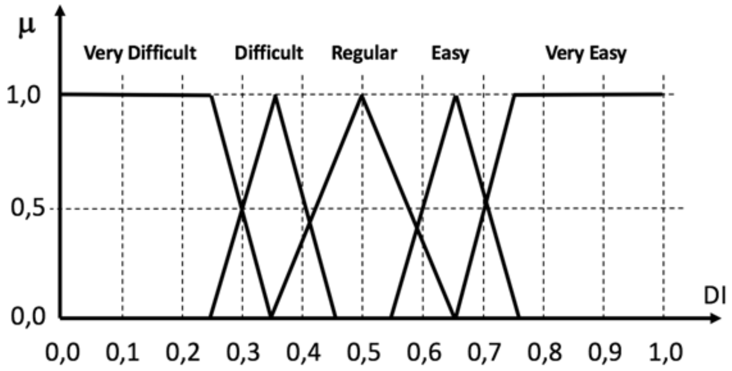

Figure 1. Fuzzy sets for difficulty index total, DI_{TOTAL}.

This means that, in case all the questions are MCQ-3 type, the penalty will be half the value of the right answered. If the number of possible answers is different in some questions, the local penalty for each different question would apply. Table 2 shows a hypothetical result of the presented example.

Table 2. Difficulty Index per question and global grade per student.

Question →	***1***	***2***	***3***	***4***	***5***	
Weight (%) →	***15***	***20***	***15***	***30***	***20***	
Student ↓						***Grade over 100 points ↓***
1 →	-7.5	0	15	30	20	***57.5***
2 →	-7.5	20	0	0	0	***12.5***
3 →	-7.5	20	0	30	-10	***32.5***
4 →	0	0	15	0	20	***35***
5 →	15	0	-7.5	0	20	***27.5***
DI →	*0.2*	*0.4*	*0.4*	*0.4*	*0.6*	

In example shown in Table 2, DI_{TOTAL} = 0.4. Thus, we can calculate the fuzzy difficulty index (FDI_{TOTAL}) with fuzzy sets shown in Figure 1. We want to highlight the use of Excel® for all calculations in this paper. After fuzzification, Table 3 shows the values of membership subsets.

Table 3. Fuzzy total difficulty index (FDITOTAL) of the questionnaire.

Fuzzy sets	Very Poor	Poor	Regular	Difficult	Very Difficult
FDI_{TOTAL}	0	0.5	0.5	0	0

In order to place the student grade in one of, for example, five categories, it is important to place his initial grade in a fuzzy environment like shown in Figure 2.

Some authors cited in [21] make a different proposal for the distribution of these subsets in order to categorize students' grades placing any grade between 0-0.2 points only in the category of 'poor' and between 0.85 and 1.0 only in the 'good' category. However, most of authors establish the limit of 'poor' between 0 and 0.1 and 'good' between 0.9 and 1.0. Continuing with the previous example, for calculate the membership to each fuzzy set, we need to obtain the equations of every subset segment, for example, for interval 0.1-0.3 (*Below Average*), the equation used is $y=5x+1{,}5$. Fuzzy grades for each student of the example are shown in table 4.

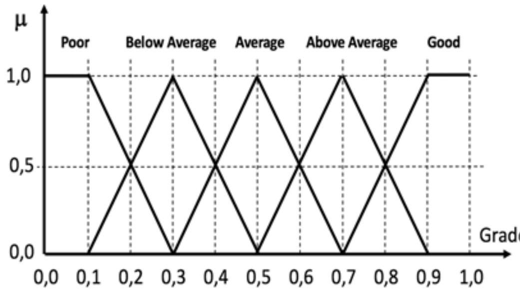

Figure 2. Fuzzy sets for grades.

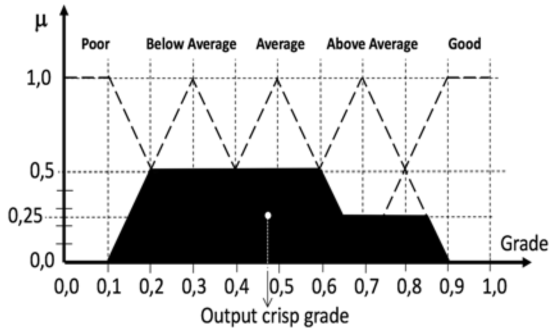

Figure 3. Output fuzzy subsets for student #4.

Table 4. Fuzzy membership grades for every student.

Student	***1***	***2***	***3***	***4***	***5***
Crisp Grade Normalized to 1	0.575	0.125	0.325	0.35	0.275
Poor	0	0.875	0	0	0.125
Below Average	0	0.125	0.875	0.75	0.875
Average	0.625	0	0.125	0.25	0
Above Average	0.375	0	0	0	0
Good	0	0	0	0	0

With this information, it is possible to make the rule base in terms of '*IF-THEN*' sentences. Consulted literature refers to different techniques for establish the rules; in this example (see Table 5), each category of output is present in five cells (it is possible to make small changes to this rule base for special cases).

Table 5. Rule base for fuzzy grades and difficulty index.

		Difficulty Index				
		Very Poor	***Poor***	***Regular***	***Difficult***	***Very Difficult***
Grades	**Poor**	Poor	Poor	B. Average	B. Average	Average
	Below Average	Poor	Poor	B. Average	Average	A. Average
	Average	Poor	B. Average	Average	A. Average	Good
	Above Average	B. Average	Average	A. Average	Good	Good
	Good	Average	A. Average	A. Average	Good	Good

Considering Mamdani inference, results of the example is listed in Table 6.

Table 6. Result of applying the base rules to each student.

	Student	***#1***	***#2***	***#3***	***#4***	***#5***
	Crisp Grade	*0.575*	*0.125*	*0.325*	*0.35*	*0.275*
Resulting fuzzy grades	Poor	0	0	0	0	0
	Below Average	0	0.5	0.5	0.5	0.5
	Average	0.5	0.125	0.5	0.5	0.5
	Above Average	0.5	0	0.125	0.25	0
	Good	0.375	0	0	0	0

If we take, for example, the student number #4, his score would be the result of calculating the center of the area resulting from the fuzzy subsets trimmed to each output value, as we can see in Figure 3.

Finally, it is necessary to return these assessment values to crisp quantities by means of one of the different defuzzification techniques that allow obtaining a unique numerical value that represents them properly in [32]. The technique of singletons [33] is the more easy and concrete way for accurate calculation. Applying singletons, the scores that the calculation would return values showed in Table 7.

Table 7. Results of defuzzified grades vs. initial obtained grades.

Student	***1***	***2***	***3***	***4***	***5***
Initial Grade Normalized to 1	0.575	0.125	0.325	0.35	0.275
Deffuzified Grade Normalized to 1	0.712	0.353	0.455	0.473	0.413

3 DISCUSSION

Some authors such as [21] rightly point out that the choice of defuzzification method depends on each context or on each specific problem, reaching the possibility of using different techniques depending on the fuzzy output subsets for each student, depending on what categories include and not use the same technique for students with grades that are *poor*, *below average*, *average*, since the same technique for all grading intervals can harm some students. This is completely true since a student with an initial score of 100 in a very easy test, could lose a grade and that would not be fair because he has guessed all the answers and should not be penalized, while a student with a poor grade in a high difficulty test, possibly know more than what the test has allowed him to manifest.

Likewise, other authors also indicate that not only is there no universal defuzzification operator but that their choice depends on factors such as the speed at which this operation should be done. For example, in the case of industrial controls in real time where computational efficiency is a very important factor or in the case of information support systems where the calculation time is not as important, as is the case of adjusting student grades. The method of extended intervals in order to defuzzify outputs is correct for the qualification in students, but considering that it will be necessary to take some correction elements to differentiate the cases in which the student may be affected by the calculation.

On the other hand, any fuzzy calculus technique is very variable since it depends on:

1) The number and geometry of linguistic subsets of each variable.
2) The number of input variables, since more variables can be used in addition to the rating and the DI_{Total}, as seen in the literature review.
3) Geometry and position of input and output fuzzy subsets.
4) Rules contained in the rules base. As it is said in [34], the rules have to be determined by expert experience and it is difficult to make a determination of a

system designed according to the fuzzy logic; that is, it cannot be estimated how the system reacts beforehand.

5) The way to make the inference, which is this case has been taken mamdani inference, but there are many other ways to do it (see, for example, [35]).
6) Defuzzification method used. They are multiple different defuzzification methods.
7) Etc.

Each change in the above factors will cause a different adjustment of results. However, at least, it is an adjustment that does not imply suppression of questions that all or no one answered correctly (with no or maximum difficulties), and should not harm anyone. In addition, it involves the concept of "justice" that each professor should establish and that, undoubtedly, it varies from professor to professor, since it is based on his personal way of understanding teaching, the fair assessment, and the corrective actions of results that are frequently doubtful.

Thus, the concept of "justice in assessment" that we apply in our case, obeys the following rules:

1) Students with an initial score of <= 10 points are considered suspended regardless of the DI_{Total} and do not considered in the correction grade.
2) Students with an initial score of >= 90 points are considered approved regardless of the di_{total} and do not considered in the correction grade.
3) Students with an initial score between 11 and 89 points, they will be compensated in function of the DI_{Total} and the corrected value will be taken.
4) If a student, after correction, obtains a negative grade, the grade will be 0 points.
5) If a student, after correction, obtains a grade higher than 100, the grade will be 100.

4 SUMMARY

This article has proposed a way of, without suppressing any question, making a readjustment of the grades based on fuzzy logic techniques. To do this, it is considered, on the one hand, the initial grade obtained by each student and, on the other, the total difficulty index of the test (overall difficulty index of an entire test, DI_{TOTAL}). With these two variables, an approximation can be made to a system of linguistic variables that allows correcting the final grades of each student based on the objective difficulty of the test and a set of rules established by the professor. This will revert to greater "justice" in students' mark system, since it will be a function of the difficulty of the test.

The above method tends to benefit students with worse grades when the difficulty of the test is high, while not penalize students who have obtained excellent grades. The MCQ tests allow to obtain the parameter DI_{TOTAL} and, with this index, to be able to establish a re-calculation of the qualification of each student.

REFERENCES

[1] Dehnad, A. Nasser, H., and Hosseini, A.F. (2014). A Comparison between three-and four-option multiple choice questions. *International Conference on Current Trends in ELT. Procedia – Social and Behavioral Sciences* 98 (2014) pp. 398-403.

[2] Rodriguez, M. C. (2005). Three-options are optimal for multiple-choice items: A metaanalysis of 80 years of research. *Educational Measurement: Issues and Practice, 24(2), pp. 3-13.* Available on-line in https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1745-3992.2005.00006.x

[3] Haladyna, T.M. and Downing, S.M. (1993). How many options is enough for a multiple-choice item? *Educational and Psychological Measurement*, *53, pp. 999-1010.*

[4] Zadeh, Lotfi A. (1965). Fuzzy sets and systems*. In J. Fox, editor, System Theory. New York: Polytechnic Press, pp. 29-39.*

[5] Ramot, D., Friedman, M., Langholz, G., and Kandel, A. (2003). Complex Fuzzy Logic. IEEE Transactions on fuzzy systems, Vol. 11, No. 4, august 2003.

[6] Biswas, R. (1995). An application of fuzzy sets in students' evaluation*. Fuzzy sets and systems pp.187-194.*

[7] Chen, S.M., and Lee, C.H. (1999). New Methods for students' evaluating using fuzzy sets. *Fuzzy Sets Systems, Elsevier. Vol. 104, No. 2, pp. 209-218.*

[8] Hamsa, H., Indiradevi, S. and Kizhakkethottam, J.J., (2016). Student academic performance prediction model using decision tree and fuzzy genetic algorithm. *Global Colloquium in Recent Advancement and Effectual Researches in Engineering, Science and Technology (RAEREST 2016). Published by Elsevier.*

[9] Wilson, E., Karr, C.L. and Freeman, L.M. (1998). Flexible, Adaptive automatic fuzzy-based grade system. *Fuzzy Information Processing Society – NAFIPS, Conference of the North American, pp. 334-338. IEEE.*

[10] Nykänen, O. (2006). Inducing Fuzzy Models for Student Classification. *Educational Technology & Society, 9 (2), pp. 223-234.*

[11] Bai, S.-M., and Chen, S.-M. (2008). Evaluating Students' Achievement using membership functions and fuzzy rules. *Expert Systems with Applications, No. 34, pp. 399-410.*

[12] Weon, S. and Kim, J. (2001). Learning achievement evaluation strategy using fuzzy membership function. *In Frontiers on Edutation Conference. 31st Annual, Vol. 1, pp. Session T3A-19. IEEE.*

[13] Deshmukh, V.B., Mangalwede, S.R. and Rao, D.H. (2017). Student Performance Evaluation Model Based on Scoring Rubric Tool for Network Analysis Subject using Fuzzy Logic. *International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT), IEEE.*

[14] Kingsy Grace, R., Manju, S., Geetha Devasena, M.S. and Krishna Kumar, V. (2017). Student Competencie Analysis: An Automated Robust Tool Using Fuzzy Logic. *International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS).* 17-18 March 2017. IEEE.

[15] Meenakshi and Nagar, P. (2015). Fuzzy Logic Based Expert System For Student's Performance Evaluation. *2nd International Conference on Computing for Sustainable Global Development (INDIACom). pp. 11-13 March 2015. IEEE.*

[16] Saleh, I. and Kim, S. (2009). A fuzzy system for evaluating students' learning achievement. *Expert Systems and Applications, An International Journal. No. 36 Elsevier.*

[17] Hameed, I.A. and Sorensen, C.G. (2010). Fuzzy Systems in Education: A More Reliable System for Student Evaluation. *Fuzzy Systems, Book, edited by Ahmad Taher Azar*

[18] Ingoley, S. and bakal, J.W. (2012b). Evaluating Student's Performance using Fuzzy Logic. *International COnference in Recent Trends in Information Technology and Computer Science (ICRTITCS). Proceedings published in International Journal of Computer Applications (IJCA).*

[19] Kavcic, A., Navia, A. and Pedraza, R. (2003). Student Modelling Based on Fuzzy Inference Mechanisms. *The IEEE Region 8 EUROCON 2003. Computer as a Tool. Ljubljana, Slovenia 22-24 Sept. 2003.*

[20] Hellendoorn, H., and Thomas, C. (1993). Defuzzification in fuzzy controllers. *Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology 1, no. 2 (1993): 109-123.*

[21] Pradeep, A. and Thomas, J. (2015). Performance Assessment for Students using Different Defuzzification Techniques. *IJIRST, International Journal for Innovative Research in Science & Technology. Vol. 2, Issue 06, November 2015.*

[22] Mamdani, E.M. (1974). Applications of fuzzy algorithms for simple dynamic plants. *Proceedings of the IEEE, 21(2), pp. 1585-1588.*

[23] Ebel, R.L. and Frisbie, D.A. (1991). Essentials of Educational Measurement. *Prentice Hall. 5th Edition.*

[24] McCowan, R.J, and McCowan, S.C. (1999). Item analysis for criterion-referenced tests. *Research Foundation of SUNY/Center for Development of Human Services.*

[25] Backhoff, E., Larrazolo, N. and Rosas, M. (2000). Nivel de dificultad y poder de discriminación del Examen de Habilidades y Conocimientos Básicos (EXHCOBA). *Revista Electrónica de Investigación Educativa*, *2* (1).

[26] Erturk, N. O., (2015). Testing your Tests: Reliability Issues of Academic English Exams. *International Journal of Psychology and Educational Studies, 2015, 2 (2), pp. 47-52.*

[27] Yadav, R.S., Singh, V. P. (2011). Modeling Academic Performance Evaluation Using Soft Computing Techniques: A Fuzzy Logic Approach*. International Journal on Computer Science and Engineering (IJCSE), ISSN: 0975-3397, Vol. 3, No. 2, February 2011*.

[28] Karandikar, R. (2010). On multiple choice tests and negative marking*. Current science. Vol. 99 Number 8, 25 october 2010.*

[29] Bush, M. (1999) Alternative marking schemes for online multiple choice tests. *7th Annual Conference on Teaching of Computing. Belfast.*

[30] Premadasa, I. G. A reappraisal of the use of multiple choice questions. *Medical teacher 15 2-3 (1993): 237-42.*

[31] Klůfa, J. (2013). Random effects in the entrance examinations. *Procedia - Social and Behavioral Sciences 106 (2013) 2197 – 2204 4th international conference on new horizons in education.*

[32] Passino, K.M. and Yurkovich, S. (1998) Fuzzy Logic. *Addison Wesley.*

[33] Bojadziev, G. y Bojadziev, M., (2007). Fuzzy Logic for Business, Finance and Management. *Singapore: World Scientific Publishing Co. Advances in Fuzzy Systems: Applications and Theory — Vol. 23, 2nd Edition, pp. 144-150.*

[34] Semerci, C. (2004). The influence of fuzzy logic theory on student's achievement. *The Turkish Online Journal of Education Technology – TOJET. April 2004, Vol. 3, Issue 2, Article 9.*

[35] Johanyák, Z.C. (2010). Survey on Five Fuzzy Inference Based Student Evaluation Methods. *Computational Intelligence in Engineering, I.J. Rudas et al. Editors. SCI 313, pp. 219-228. Springer-Verlag.*